Claim 20 is rejected under 35 USC §103(a) as being unpatentable over *Pirolli* and *Prasad* in view of U.S. Patent No. 6,128,606, herein *Bengio*.

Claims 21 – 25 are rejected under 35 USC §103(a) as being unpatentable over *Pirolli* and *Prasad* in view of U.S. Patent No. 6,389,436, herein *Chakrabarti*.

INDEPENDENT CLAIMS (CLAIMS 1 AND 34)

Pirolli

The Office Action alleges (paper #9, page 3, the first sentence of the last paragraph),

> Pirolli discloses determining how strongly each document correpondes to each of the categories...(emphasis added).

However, the preamble of claim 1 of *Pirolli* is

> A system for identifying documents relevant to a focus document...

Also, the field of invention states,

> The present invention is related to the field of analysis and design of linked collections of documents, and in particular to predicting documents of relevance to a focus document.

In contrast claim 1, lines 6 and 7, and claim 34, lines 7 and 8, recite

> determining how strongly each document of said ***plurality of documents*** corresponds to each of said ***plurality of categories***...(emphasis added).

Thus, *Pirolli* discloses a system for determining how relevant documents are to "a focus" (i.e., a single focus), while claim 1 recites a method for determining the relevance of a "plurality" of documents to a "set" of categories. Although the Office Action cites, column 8, lines 8-47, there is no recitation in this passage of categorizing a group of document into a group of categories. Instead, column 8, lines 8-47 refer to

> ...predicting degree of category membership for each page at a web locality.

Thus, claims 1 and 34 describe a group process in which a group of documents are categorized by comparing them with a group of documents, whereas *Pirolli* does not disclose such a group process. A significance of this difference is that although claims 1 and 34 are not limited to optimizing an objective function, claims 1 and 34 are generic to being used for optimizing an objective function. When optimizing an objective function according to the method disclosed a group of documents are categorized into a group of categories. The process of *Pirolli*, however, is not generic to optimizing an objective function in this manner. The Examiner has not yet addressed this difference.

The Office Action further states (paper #9, page 4, the first full paragraph),

> Further, Pirolli discloses the sets of determining similarity performed using a matrix representing document similarity that is derived by combining two or more measures of document similarity. (Pirolli, col. 11, lines 36-39: "An activation network can be represented as a graph defined by matrix R, where each off-diagonal element $R_{i,j}$ contains the strength of association between nodes i and j, and the diagonal contains zeros."; col. 8, lines 8-13: "In order to perform categorizations each Web page at the Web locality is represented by a vector of features constructed from the above topology, meta-information, usage statistics and paths, and text similarities. These Web page vectors are collected into a matrix. Such a matrix is illustrated in FIG. 5.")

However, although column 11, lines 36-39, referenced by the Office Action, describe a matrix, this matrix is for the purpose that (column 11, line 36) "An activation network can be represented...." In other words the matrix of column 11 lines 36-39 are for representing an activation network and not for categorizing. Consequently, column 11, lines 36-39 are contained within a section entitled (column 11, line 35), "Activation." Regarding the activation, *Pirolli* state (column 2, lines 3-8),

> The system provides for (a) categorization based on feature vectors that characterize individual page information and (b) prediction of need (or relevance) of other Web pages with respect to a particular context, which could be a particular page or set of pages, using a spreading activation technique.

Thus, the activation is used for the prediction of the need for the relevance of other Web pages with respect to a particular context, and not with respect to categorizing. Therefore, the matrix of column 11, lines 36–39, is not only for activation and not for the categorization recited in claims 1 and 34, but is also unrelated to the categorization discussed in column 8, lines 8-13, contrary to the implications of the Final Office Action (paper #11).

Regarding column 8, lines 8-13 (cited by the Office Action, paper #9), and the step of "determining" of the last paragraph of claims 1 and 34, it is not clear to what extent the Office Action is relying on the column 8, lines 8-13, and to what extent the Office Action is relying on *Prasad*. Specifically, the "similarity" referred to in the last paragraph of claims 1 and 34 refers to the similarity between each document and a set of training documents, which the Examiner acknowledges *Pirolli* does not disclose. (Yet, it is possible to misinterpret the Office Action as alleging that all of the last paragraph of claim 1 is disclosed by *Pirolli*) However, the difference between *Pirolli* and the last paragraph of claim 1 is yet deeper, as perhaps the Examiner would agree. Specifically, the matrix of *Pirolli* is made by collecting together each web page's vector of features into one matrix for only one web locality. In contrast, claims 1 and 34 recite that the matrix is made by combining two or more measures of document similarity, where as described in the previous paragraph of claims 1 and 34, the similarity is between two documents from two different groups. It is not clear if these two matrices are the same thing. If they are the same thing, the Office Action does not explain why they are the

same thing. If they are not the same thing, the Office Action does not explain why the difference is allegedly obvious.

Further, the similarity matrix of claims 1 and 34 is for determining the similarity of a plurality of documents to a plurality of categories as specified in the second to last paragraph of claim 1 and 34. Thus, even if otherwise the similarity matrix of claims 1 and 34 has a similar content to that of *Pirolli*, the matrix of claims 1 and 34 in at least one sense contains additional content in that it relates to a plurality of categories and a plurality of documents to be categorized and not to only one web locality or focus (and a plurality of web pages related to that web locality or focus). Thus, in contrast to the matrix of *Pirolli*, the matrix of claim 1 and 34 is well suited for the optimization process disclosed, without adding other matrices. Perhaps the Examiner would agree with this, however, clarification is respectfully requested.

As explained in the response filed April 23, 2003, the Applicant admits that *Pirolli* teaches (1) to categorize a set of documents, in the form of pages, according to "classification characteristics," and (2) to determine textual similarity between documents to categorize a document. However, the Applicant is not attempting to claim only these features. Rather, the Applicant is claiming use of the similarity between documents in a group of documents and a particular set of documents (i.e., a training set), which have been established as belonging to a group of categories, to determine the correspondence between the group of documents and the group of categories.

As also explained in the response filed April 23, 2003, *Pirolli* teaches that documents are categorized into functional categories (which are designed by a person). A number of characteristics are used to classify documents. Only one of these characteristics are based on similarity between a document and a particular set of

documents, while claims 1 and 34 recites a similarity matrix for representing document similarity that combines at least two measures of document similarity. The one characteristic of similarity in *Pirolli*'s matrix is csim; "csim, [is] the textual similarity of the item to it's children based upon previous SCA calculation (column 508)," which is a single number rather than a matrix representing document similarity derived by combining two or more measures of document similarity.

The Final Office Action (paper #11, the first paragraph of page 4) states,

> Applicants next argue (page 6) that Pirolli "seems to teach against such a feature because of the types of functional categories it discloses." However, just because Pirolli discloses different kinds of categories and categorization than what is claimed by applicants does not mean that Pirolli teaches against their claimed invention.

The Final Office Action is apparently referring to statements in the response, such as

> In fact, *Pirolli* seems to teach against such a feature because of the types of functional categories it discloses. For example, head node is a category which includes documents in which text similarity between the documents in this category is of little relevance. Examples of a set of documents that could be established in this category are Yahoo's home page, Google's home page, and the USPTO home page. It would seem that text similarity between these pages and another page would have very little relevance to whether the other page is a home page.

The Final Office Action apparently agrees that these other categories are not those recited in the claims, and (as is now apparent) was not relying upon them in making the rejection. Consequently, the only issue is whether the csim can be relied upon in a rejection under 35 USC §103. Is that the intent of the Office Action? Clarification is respectfully requested.

As pointed out in the response filed April 23, 2003, *Pirolli* further teaches that text similarity is used to determine whether a page belongs to the category of head page (e.g., home page) (col. 9, lines 14 – 24).

> For Head Nodes (classification criteria 601), being the first pages of a collection of documents with like content, it is expected that such pages will have high text similarity between itself and its children, and would have a high average depth of its children, and that it would be more likely to be an entry point based upon actual user navigation patterns.

Thus, at best, *Pirolli* teaches that text similarity between a page and the children of the page is used to determine the correspondence between the page and the category of home page. However, this is not a category to which the set of children have been established as belonging. The claims, on the other hand, require the feature of using similarity between a group of documents and a particular set of documents established as belonging to a category to determine the correspondence between the group of documents and the category. However, as indicated by the Final Office Action, perhaps the Office Action was not relying on the Head Nodes category the rejection.

It is not clear what precisely, the Office Action is relying on within *Pirolli*, because *Pirolli* has categories, documents, web pages, and web localities. However, the matrices of *Prirolli* (e.g., FIG. 5) relate documents to the web localities and not to categories or else they relate documents to documents, but are used for activation and not for categorization, in contrast the matrices of claims 1 and 34, which relate documents from one group to documents in another group used for categorization.

Any ambiguities in the Office Action, however, are at least in-part due to the ambiguities of *Pirolli*. Although *Pirolli* may have provided an adequate disclosure for supporting their claims, *Pirolli* contains ambiguities as to which documents are categorized into which categories, and for what purpose, which at least diminishes its

usefulness as a reference in a rejection under 35 USC §103. Somehow *Pirolli*'s categorization of documents is related to the determining the relevance of web pages at a web locality to a search query, but the precise nature of the categorization and how it is used or related to the relevance of web pages is unclear.

Pirolli in view of Prasad

As a result of ambiguities in *Pirolli*, it is not clear that one of ordinary skill in the art would be motivated to modify the categorization of *Pirolli* using the categorization of *Prasad*, without knowing the precise purpose of *Pirolli*'s categorization.

The argument to which the first paragraph on page 4 of the Final Office Action (paper #11) refers are also in part based on column 8, lines 35-37 of *Pirolli* which state,

> categories are designed by someone (application designer, webmaster, end user), in contrast to being automatically induced.

This statement of *Pirolli* is not merely a recitation of kinds of categories and categorizations, but evidence that *Pirolli* recognize their process for determining rules could be automated (and which the Final Office Action presumably is alleging would be obvious to modify by "automating" it according to *Prasad*), and *Pirolli* teach to not automate the process of determining rules, thereby precluding automating their process using the teachings of *Prasad* in a rejection under 35 USC §103. This point, mentioned in the response filed on April 23, 2003 (page 6, the second to last paragraph) was not explicitly addressed by the Final Office Action.

To elaborate on this point, *Pirolli* imply that the categories should be made by human beings and not made automatically. As explained in the response filed June 25, 2003, *Pirolli* is addressing the problem of the "sluggishness" associated with prior art

searching techniques (see column 1, lines 25-28). Consequently, the reason *Pirolli* like the use of rules made by human beings is

> Based on category membership, a user may ***quickly*** predict the functionality of an element. For instance, in the everyday world, identifying something as a "chair" enables the quick prediction that an object can be sat on... (emphasis added, column 8, lines 53-55).

In other words, an important point being made here is that a reference about a chair may not mention anything about sitting, but by using rules one can nonetheless quickly make an association between the chair and sitting. Similarly, using rules one can make an association between a document and how to categorize it, even though the document may not explicitly mention anything about many of its attributes. However, one of ordinary skill in the art would expect that such an advantage would be lost were one to use a bunch of training documents to establish the rules because the rules established from the training are unlikely to include concepts that are not explicitly discussed in the training documents and because using training documents increases the time to establish the rule.

Therefore, one of ordinary skill in the art would be inclined not slow down the categorizing process by using the more limited rules derived from training documents of *Prasad.* In this sense the *Prasad*'s use of the training documents runs contrary to at least one of the principals upon which *Pirolli* 's system is based, which is not permitted in a rejection under 35 USC §103, (see MPEP 2143.01, p. 2100-127, the right column, entitled, "THE PROPOSED MODIFICATION CANNOT CHANGE THE PRINCIPLE OF OPERATION OF A REFERENCE," which cites *In re Ratti*, 270 F.2d 810, 123 USPQ 349 (CCPA 1959)).

Thus, as explained in the response filed April 23, 2003, *Prasad* fails to teach the claimed feature of using similarity between a document and another set of documents

established as belonging to a category to determine the correspondence between the document and the category. Presumably, the Office Action has equated a document as claimed to a document at a data source and a training set as claimed to a sample of documents from a data source. Even if the training set taught by *Prasad* can be equated to the training set claimed, *Prasad* nevertheless fails to teach the claimed feature.

The Final Office Action states (page 4, the second paragraph),

> The examiner disagrees with applicants' characterization of *Prasad* inasmuch as the rule induction taught by *Prasad* is used to classify documents, i.e., determine their similarity to a category. (*Prasad*, col. 4, lines 3-16.)

However, column 4, lines 3-16 of *Prasad* state

> As a solution to providing an automatic and optimal selection of desired data sources for user queries, a form of supervised machine learning called "Rule Induction" generates a model for classifying the sources **20** for query searching. The model is then used for predicting the top "N" sources most likely to contain documents that satisfy a user's query. As an overview, "Rule Induction" takes a sample set of documents called a training set and derives "Disjunctive Normal Form Rules" representative of the model which is descriptive of the data sources **20**. "Rule Induction" is often the preferred approach to classification modeling and prediction due to the enhanced capability and interpretability of decision rules in responding queries.

The Applicants respectfully submit that contrary to the implications of the Final Office Action, column 4, lines 3-16 of *Prasad* never suggest "the rule classification taught by *Prasad* is used to classify documents, i.e. determine their similarity to a category." Instead, first " 'Rule Induction' generates a model for classifying sources 20 for query searching." This is done by (column 3, lines 19-23)

> A prior art algorithm is used to recognize patterns in the sets of samples to distinguish one source from another and generate a set of Disjunctive Normal Form (DNF) Rules, as a model, representing each source.

Alternatively, as stated in column 4, lines 10-13,

> "Rule Induction" takes a sample set of documents called a training set and dervies "Disjunctive Normal Form Rules" representative of the model which is descriptive of the data sources 20...

In other words, the sources 20 are the "categories" into which the documents are already located, and in this sense preclassified, and rules are derived for determining the common characteristics of the documents that distinguish them from the documents of other sources. For example (column 3, lines 16-19),

> A dictionary is created to define features and attributes representing individual sources. All documents are transformed into a set of samples comprising a feature, a word or phrase and a source name used in the dictionary.

After deriving rules for the sources (column 4, lines 7-9),

> The model is then used for predicting the top "N" sources most likely to contain documents that satisfy a user's query.

Thus, column 4, lines 3-16, disclose using documents in a source to derive characteristics of a source for formulating rules that are used for finding which source is most likely to contain a document that meets a search query. Column 4, lines 3-16, do not disclose classifying new documents by comparing them to other documents. Instead *Prasad* teaches that rule induction is applied to the training set to generate rules that are used to determine which source to direct queries (col. 3, line 66 – col. 4, line 16). While *Prasad* teaches that training sets are used as input for rule induction, no teaching in *Prasad* suggests training sets are used determine the correspondence between a document and the category to which the training set belongs by determining the similarity between the document and the training set.

As also explained in the response filed June 25, 2003, *Prasad* is attempting to determine which source to retrieve documents from, while *Pirolli* is attempting to

categorize documents found. In this sense these two documents may not even be related art. Cf. MPEP2141.01(a) p.2100-118, which cites *In re Clay*, 966 F.2d 656, 23 USPQ2d 1058 (Fed. Cir. 1992) and emphasizes the difference between "storage" and "extraction" as significant in determining a reference to be non-analogous (the difference between storage and extraction is conceptually very similar to the difference between categorizing search results and identifying sources where to search).

Deciding which source to retrieve documents from is analogous to deciding whether to use Lexis', INSPEC's, or Dialog's databases to find a document. The source where a document is found is not necessarily a useful category for classifying search results. The difference between these sources is not ordinarily associated with differences between two categories in which documents found are likely to be classified into.

Further, the claims require that the training documents be already categorized into the categories. In *Prasad*, it would appear that the training documents happen to already be in the sources before the search began with no effort on the part the developer to categorize the documents. While the claims do not necessarily require effort or a pre-categorization step on the part of a developer, the effort of pre-classification typically required in finding training documents for categorizing (not necessary when deciding on sources) seemingly would have deterred one of ordinary skill from using training documents when categorizing, and would have caused one of ordinary skill in the art to think of these two activities as unrelated distinct processes. Were one of ordinary skill to have combined *Prasad* and *Pirolli*, it would have been to use *Prasad*'s training documents to decide on which source to take the documents from and not in categorizing and ranking the documents later found. Thus, it would seem unlikely that one of ordinary

skill in the art would look to a reference on where to search, to solve a problem about how to categorize search results.

In view of the deficiencies pointed out above, in order to expedite the prosecution, the remaining deficiencies with the references relied upon in rejecting claims 1 and 34 will not be discussed at this time.

DEPENDENT CLAIMS

Each of dependent claims 2-33 contains features that are independently patentable over the prior art. Some examples are discussed below.

Pirolli and Prasad in view Bengio.

Regarding claim 20, the Office Action stated (paper # 9, the first full sentence of page, 4),

> Because claim 20 is directed to a similar invention, it would have been obvious to one of ordinary skill in the art to have combined Pirolli, *Prasad*, and Bengio to implement the optimization of an objective function.

However, MPEP 2143.01, p. 2100-126, states

> **FACT THAT REFERENCES CAN BE COMBINED OR MODIFIED IS NOT SUFFICIENT TO ESTABLISH *PRIMA FACIE* OBVIOUSNESS**
> The mere fact that references can be combined or modified does not render the resultant combination obvious unless the prior art also suggests the desirability of the combination. *In re Mills*, 916 F.2d 680, 16 USPQ2d 1430 (Fed. Cir. 1990)

Further, MPEP 2143.01, p. 2100-126, states

> **FACT THAT THE CLAIMED INVENTION IS WITHIN THE CAPABILITIES OF ONE OF ORDINARY SKILL IN THE ART IS NOT SUFFICIENT BY ITSELF TO ESTABLISH *PRIMA FACIE* OBVIOUSNESS**

> A statement that modifications of the prior art to meet the claimed invention would have been " 'well within the ordinary skill of the art at the time the claimed invention was made' " because the references relied upon teach that all aspects of the claimed invention were individually known in the art is not sufficient to establish a *prima facie* case of obviousness without some objective reason to combine the teachingsof the references. *Ex parte Levengood*, 28 USPQ2d 1300 (Bd. Pat. App. & Inter. 1993). See also *In re Kotzab,* 217 F.3d 1365, 1371, 55 USPQ2d 1313, 1318 (Fed. Cir. 2000)

Although being in related fields of endeavor is a prerequisite to being able to combine references in a rejection under 35 USC §103, it logically follows from the above principles of MPEP 2143.01, that just because two references are in the same field or in the Office Action's terminology, "directed to a similar invention," does not in-and-of-itself establish a motivation to combine the two references or otherwise make the combination obvious to one of ordinary skill in the art.

Pirolli and Prasad in view of Chakrabarti.

Regarding claim 22, the Office Action stated (page 14, the second to last paragraph)

> However, given that a growth function is one which by definition stabilizes in a finite number of steps, it would have been obvious for one of ordinary skill in the art to have extended the combination of Priolli, *Prasad*, and Chakrabarti to repeatedly apply a growth transformation

However, this statement is apparently based on the Applicant's specification, which states (page 27, lines 13-15),

> An advantage of Growth Transformation is that it guarantees the monotonic increase of the objective function in a finite step, instead of infinitesimal step as most gradient algorithms do.

Yet, MPEP 2145 (X)(A) refers impermissible hindsight as " 'knowledge gleaned only from applicant's disclosure,' " and cites *In re McLaughlin* 443 F.2d 1392, 1395, 170

USPQ 209, 212 (CCPA 1971). Although the application admits that growth functions are old, and "have been applied in the past to maximizing the constrained polynomial objective function, and re-estimation of statistical model parameters of hidden Markov models." There is no admission of growth functions being applied to categorizing documents. Although growth functions were known in other arts, it is not clear whether at the time of the invention, one of ordinary skill in the art pertinent to claim 22 would have even known what a growth function is, or known what its advantages are.

In view of the patentable features pointed out above, in order to expedite the prosecution the remaining differences between the prior art and the various dependent claims will not be discussed at this time.

For the reasons set forth above, the Applicants respectfully submit that all pending claims are patentable over the art of record, including the art cited but not applied. Accordingly, allowance of all claims is hereby respectfully solicited.

Respectfully submitted,

HICKMAN PALERMO TRUONG & BECKER LLP

Dated: August 15, 2003

David Lewis
Reg. No. 33,101

1600 Willow Street
San Jose, CA 95125
Telephone No.: (408) 414-1080 ext.206
Facsimile No.: (408) 414-1076

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as first-class mail in an envelope addressed to: Mail Stop No Fee -- AF, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

on August 15, 2003 by Teresa Austin